DOCUMENT RESUME

ED 048 363 | TM 000 441

AUTHOR	Forster, Fred
TITLE	The Generalized Johnson-Neyman Procedures: An Approach to Covariate Adjustment and Interaction Analysis.
PUB DATE	Feb 71
NOTE	22p.; Paper presented at the Annual Meeting of the American Educational Research Association, New York, New York, February 1971
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*Analysis of Covariance, Classification, Computer Programs, Hypothesis Testing, Individual Characteristics, *Interaction, Mathematical Models, *Multiple Regression Analysis, Predictive Measurement, *Reliability, *Research Methodology, Tests of Significance
IDENTIFIERS	*Johnson Neyman Procedures

ABSTRACT

Statistical methods are described for diagnosing and treating three important problems in covariate tests of significance: curvilinearity, covariable effectiveness, and treatment-covariable interaction. Six major assumptions, prerequisites for covariate procedure, are discussed in detail: (1) normal distribution, (2) homogeneity of variances, (3) covariable-group independence, (4) reliability, (5) linearity, and (6) homogeneity of regression. A generalization of the Johnson-Neyman tests of significance (originally developed for two groups and two covariables, but frequently ignored by many critics of analysis of covariance) to cover any number of groups and covariables is presented. The procedure is viewed as a powerful tool for measuring the relationship between learner characteristics and teaching strategies when the regression slopes are not homogeneous. Aided by computer technology, it is proposed as a relatively easy method for classification of students by their individual needs as well as by the characteristics of teaching methods. (CK)

THE GENERALIZED JOHNSON-NEYMAN PROCEDURES: AN APPROACH TO COVARIATE ADJUSTMENT AND INTERACTION ANALYSIS

Fred Forster
Northwestern University

THE GENERALIZED JOHNSON-NEYMAN PROCEDURES: AN APPROACH TO COVARIATE ADJUSTMENT AND INTERACTION ANALYSIS

The purpose of this paper is to describe statistical methods for diagnosing and treating three important problems in covariate tests of significance--curvilinearity, covariable effectiveness and treatment-covariable interaction[1].

Assumptions Supporting Covariate Analyses

Some recent articles have described the fundamental assumptions of most covariate models[2], but they neglected to explore the relative importance (or unimportance) of each assumption especially when applied to quasi-experimental designs[3]. Exploring the relative importance of the assumptions is crucial because the quasi-experimental design places added demands and stresses upon the analysis. While the reason for employing a covariate test in experimental designs is to improve statistical precision (rarely of critical concern), the reason for employing a covariate test in quasi-experimental designs is to adjust for unknown group biases due to non-random assignment (always of critical concern).

In brief, the major assumptions behind a coveriate procedure are the following:

1. Normality--criterion and covariables are assumed to be normally distributed.

2. Homogeneity of variances--the variances of criterion and covariables are assumed not to differ among groups.

3. Covariable-group independence--the groups are assumed to be drawn from a single underlying population, and each group reflects the population covariable-dependent variable relationship.

4. Reliability--the covariables are assumed to be free of measurement error.

5. Linearity--the covariables are assumed to be linearly related to the criterion.

6. Homogeneity of regression--the group regression equations are assumed to be independent of treatments[4].

Each of these assumptions will now be discussed in detail.

Normality and Homogeneity of Variance

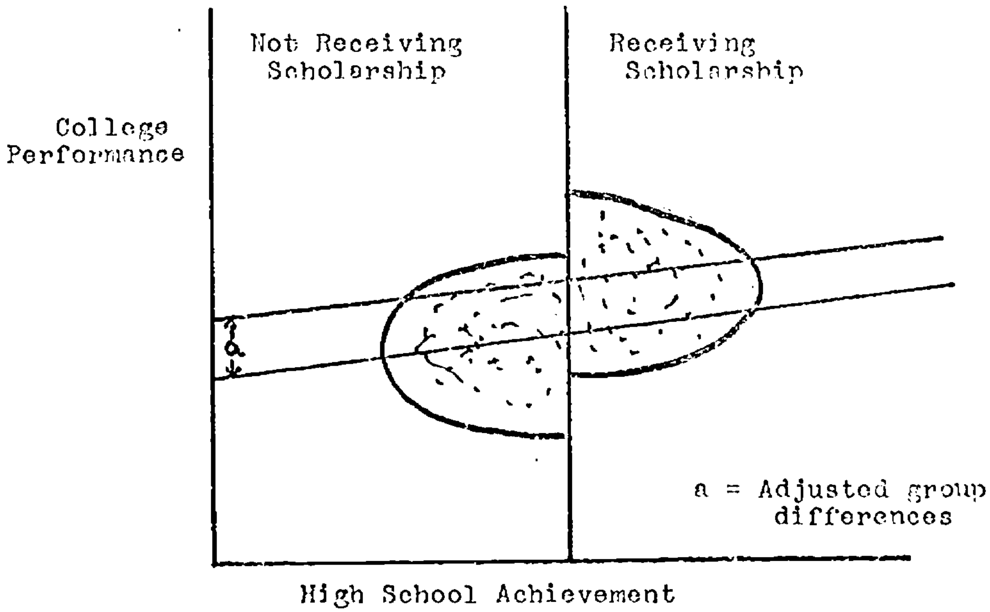
Happily it has been shown that Analysis of Covariance (hereafter called ANCOVA) is robust to violations of normality and homogeneity of variance for experiments and quasi-experiments unless the deviations among groups are bizarre, and therefore neither of those assumptions need be of critical concern[5,6].

Covariable-Group Independence

Covariable-group independence is a sine qua non for quasi-experimental covariate analyses[7]. A situation which may seem to refute this assumption is the assignment of subjects to groups entirely on the basis of covariable scores[8].

FIGURE 1

THE ARBITRARILY PARTITIONED SINGLE POPULATION



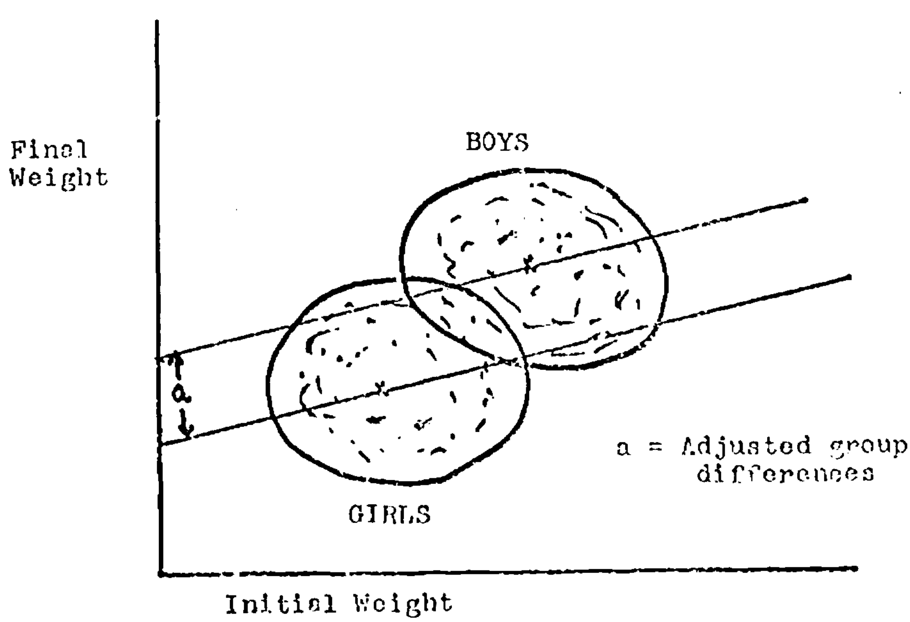
As shown in Figure 1, this design is most appropriate when a scarce commodity (like academic scholarships) is dispensed

on the basis of previous achievement. The differential performance represented by a is a measure of the extra effect of the scholarship beyond the impact of the covariable. The power of this analysis depends on the amount of data near the cut points and, consequently, on the continuity between the group regression lines. For that reason the example of Figure 1 is an ideal situation since the data is most plentiful at the cut point. The success of this analysis depends on the fact that the groups represent distinct segments of a single normally distributed population so that the assumption is not violated (only stretched a little).

Figure 2 illustrates a hypothetical example of Lord's

FIGURE 2

LORD'S ANCOVA PARADOX



about the problems which can result when this assumption is violated[9]. A researcher attempting to contrast the effect of two diets makes his initial weight measurements and then assigns one diet to a boys dorm and the other to a girls dorm. Although the initial and final weight distributions of both groups are identical, ANCOVA would lead to the conclusion that boys gained more than girls. Skipping the statistical artifacts, it is clear that criterion and covariable have been hopelessly confounded with group membership (boys as a group outweigh girls). The point of this discussion is that in a quasi-experiment it must make sense to equate the groups, i.e. the groups must represent the same basic population.

Reliability

Although errors in measurement are important, they are usually beyond the control of the researcher. It has been shown that covariable reliability levels above .75 are sufficient for most situations although a reliability estimate can always be used to improve the precision of the analysis[10]. As with all the assumptions that follow, reliability is far more critical an issue for quasi-experimental designs than for true experiments.

Curvilinearity

In principle, curvilinearity should not be a problem since all covariate models can be easily extended to include nonlinear terms. In practice, the problem of detecting curvilinearity and then systematically testing alternative regression models requires a good deal of effort. The much cited practice of "eyeballing" scatterplots though intuitively appealing is just not reliable enough for most analytic purposes. A more effective detection method utilizes tests for fit and departure from fit[11]. If these tests are incorporated into a stepwise model (say an increasing polynomial) then terms can be added until the fit is most significant and the departure from fit is not significant.

Figure 3 illustrates this stepwise analysis applied to two variables of teacher performance where the best fit has been identified as a cubic polynomial. After the best

FIGURE 3

A STEPWISE TEST FOR CURVILINEARITY

Step 1: $y = b_o + b_1x$; $r = -.261$

ANOVA Table

Source	df	Sum of Sqs.	Mean Sq.	F	p
Regression	1	52.445	52.445	2.713	.119
Departure	12	445.079	37.090	1.919	.122
Errors	14	270.583	19.327		
Total	27	768.107			

FIGURE 3--Continued

Step 2: $y = b_0 + b_1x + b_2x^2$; Eta = .339

ANOVA Table

Source	df	Sum of Sqs.	Mean Sq.	F	p
Regression	2	88.177	44.089	2.281	.138
Departure	11	409.347	37.213	1.925	.124
Errors	14	270.583	19.327		
Total	27	768.107			

Step 3: $y = b_0 + b_1x + b_2x^2 + b_3x^3$; Eta = .479

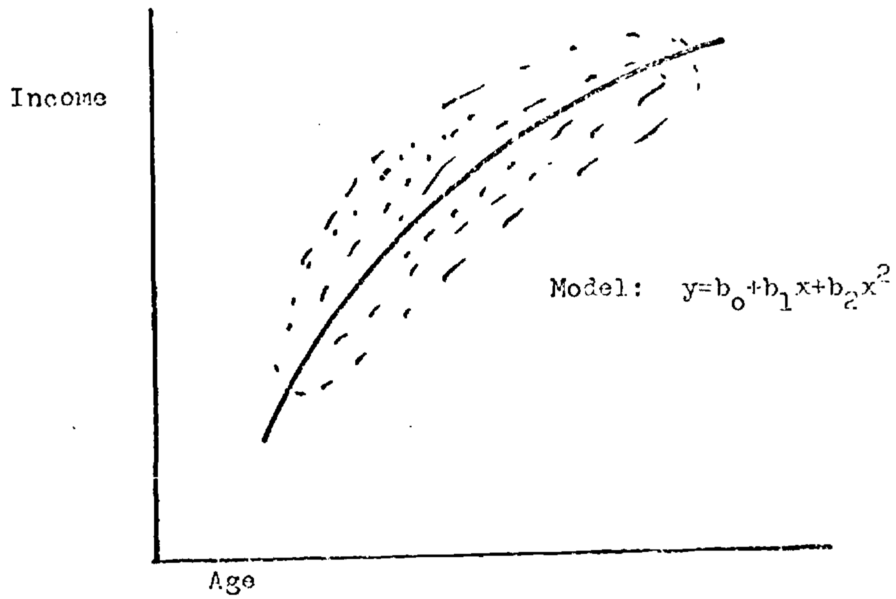
ANOVA Table

Source	df	Sum of Sqs.	Mean Sq.	F	p
Regression	3	176.035	58.678	3.036	.064
Departure	10	321.489	32.149	1.663	.186
Errors	14	270.583	19.327		
Total	27	768.107			

statistical model has been identified in this way, it is important to "think through" the relationship and square it with the theoretical framework of the study. The justification of a model chosen by an arbitrary procedure like this lies in its shape, not its order. For example, although the fit between age and income in Figure 4 is clearly curvilinear, the quadratic equation should be viewed as only a good first approximation to the true relationship[12].

FIGURE 4

EXAMPLE OF A CURVILINEAR FIT



Sufficiency and Efficiency

Although sufficiency and efficiency are not assumptions made by covariate analysis they pose special problems for quasi-experimental designs. Sufficiency essentially depends on the ability of the researcher to identify covariables which account for <u>every major bias</u> which exists because of nonrandom assignment to groups. Efficiency, on the other hand, is important because covariate methods are unusually susceptible to inaccuracies due to redundancy among the covariables. In

fact, if two perfectly correlated covariables are used the calculations will completely break down[13]. Once a sufficient set of covariables is available two procedures can be used to locate and remove redundancy. First, all covariable pairs can be prescreened to identify pairs where the gain from the second covariable is less than a specified amount (say 10 %) of its zero order contribution. When a problem pair is identified, the least meaningful covariable can be deleted or a new variable can be created combining both variables[14]. Second, a stepwise regression can be performed groupwise to identify the efficient set for each group. To be consistent in this analysis all terms of a curvilinear covariable should be added in a single step. Finally, the union of the efficient covariable sets for each group is taken as the covariable set for the analysis. As with curvilinearity, the final set of covariables must be related to the theoretical model to insure that results will be interpretable and meaningful.

Homogeneity of Regression

When the criteria described above have been satisfied, the threat of nonhomogeneity of regression still remains. Although ignored by many critics of ANCOVA, a powerful series of tests was developed by Palmer O. Johnson and Jersey Neyman to detect treatment-regression interactions and perform the

required tests of significance for two groups and two covariables[15]. What follows is the mathematical derivation of the author's generalization of these tests to cover any number of covariables and groups.

Assume k groups, n efficient covariables (x_1, x_2, ..., x_n) and a group regression vector $\underline{B}_i = (b_{oi}, b_{ii}, \ldots, b_{ni})$[16]. A set of four useful hypotheses can be established as follows:

H_a: $\underline{B}_1 \neq \underline{B}_2 \neq \ldots \neq \underline{B}_k$; each group has a unique regression vector.

H_1: $\underline{B}_1 = \underline{B}_2 = \ldots = \underline{B}_k$; all groups have a common regression vector.

H_2: $\underline{A}_1 = \underline{A}_2 = \ldots = \underline{A}_k$, $b_1 \neq b_2 \neq \ldots \neq b_k$; all groups have a common within group regression and different group means.

H_3: $\underline{B}_1 \neq \underline{B}_2 \neq \ldots \neq \underline{B}_k$, $\underline{X}\,\underline{B}_1 = \underline{X}\,\underline{B}_2 = \ldots = \underline{X}\,\underline{B}_k$; each group has a unique regression vector and the groups do not differ at the point $\underline{X}$.

Then the following set of powerful tests can be employed:

Test 1: Are there any significant differences among groups (H_1 vs H_a)?

Calculations: Proceeding from the general linear model $Y = \underline{X}\,\underline{B}_i + e$ where $\underline{X} = (1, X_1, \ldots, X_n)$ and $\underline{B} = (b_{oi}, b_{ii}, \ldots, b_{ni})$ for each of the k groups, then under H_a the maximum likelihood estimate $\hat{\underline{B}}_i = \underline{R}_i \underline{T}_i^{-1}$ where $\underline{R}_i = \sum_{j=1}^{N_i} \underline{X}'_{ij} \underline{X}_{ij}$; and the sum of squares deviation from H_a is

$$S_a^2 = \sum_{i=1}^{k} \sum_{j=1}^{N_i} Y_{ij}^2 - \sum_{i=1}^{k} \underline{R}_i \hat{\underline{B}}_i$$

where S_a^2 is distributed $\chi^2\left[\sum_{i=1}^{k} N_i - k(n+1)\right]$.

Under H_1 the maximum likelihood estimate

$\underline{\hat{B}} = \underline{R}\,\underline{T}^{-1}$ where $\underline{R} = \sum_{i=1}^{k} \underline{R}_i$ and $\underline{T} = \sum_{i=1}^{k} \underline{T}_i$ and

the sum of squares deviation from H_1 is

$S_1^2 = \sum_{i=1}^{k}\sum_{j=1}^{N_i} Y_{ij}^2 - \underline{R}\,\underline{\hat{B}}$. $\Delta S_1^2 = S_1^2 - S_a^2$ and

ΔS_1^2 is distributed as $\chi^2[(k-1)(n+1)]$. ΔS_1^2 represents the <u>increase in the sum of squares deviations due to H_1</u>. H_1 vs H_a can be tested by forming

$$F_1 = \frac{\Delta S_1^2/df_1}{S_a^2/df_a}$$

where $df_1 = (k-1)(n+1)$ and $df_a = \sum_{i=1}^{k} N_i - k(n+1)$

If F_1 is statistically signigicant, then it will be fruitful to proceed with the analysis. If F_1 is not statistically significant the analysis can be terminated since there are <u>no signigicant differences in group means or regressions</u>.

Test 2: <u>Assuming differences among group means, are there any significant differences among the covariable coefficients among groups</u> (H_2 vs H_a)? This test is commonly referred to as the <u>test for homogeneity of regression</u>.

Calculations: H_2 requires a partitioning of the group

$\underline{B}_i$ vectors into the group mean b_{oi} and a common within group regression $\underline{A}$, i.e. $\underline{B}_i = (b_{oi}, \underline{A})$. Under H_2 the maximum likelihood estimate of $\underline{B}_H = (\hat{b}_{oi}, \ldots, \hat{b}_{ok}, \hat{\underline{A}})$ is $\underline{B}_H = \underline{R}_H \underline{T}_H^{-1}$ where

$\underline{Z} = (X_1, X_2, \ldots, X_n)$;

$$\underline{R}_H = \left(\sum_{j=1}^{N_1} Y_{ij}, \ \vdots\vdots\vdots \ \sum_{j=1}^{N_k} Y_{ij}, \ \sum_{i=1}^{k} \sum_{j=1}^{N_i} \underline{Z}'_{ij} \underline{Z}_{ij} \right);$$

$$\underline{T}_H = \left[\begin{array}{cccc|c} N_1 & 0 & \ldots & 0 & \sum_{j=1}^{N_1} \underline{Z}_{ij} \\ 0 & N_2 & \ldots & 0 & \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \ldots & N_k & \sum_{j=1}^{N_k} \underline{Z}_{ij} \\ \hline \sum_{j=1}^{N_1} \underline{Z}'_{ij} & \vdots\vdots\vdots & & \sum_{j=1}^{N_k} \underline{Z}'_{ij} & \sum_{i=1}^{k} \sum_{j=1}^{N_i} \underline{Z}'_{ij} \underline{Z}_{ij} \end{array} \right].$$

Likewise the sum of squares deviation due to H_2 is $S_2^2 = \sum_{i=1}^{k} \sum_{j=1}^{N_i} Y_{ij}^2 - \underline{R}_H \hat{\underline{B}}_H$. $\Delta S_2^2 = S_2^2 - S_a^2$ where

ΔS_2^2 is distributed $\chi^2[n(k-1)]$, and ΔS_2^2

represents the gain in sum of squares deviations due to H_2. H_2 vs H_a can be tested by

$$F_2 = \frac{S_2^2/df_2}{S_a^2/df_a}$$

where $df_2 = n(k - 1)$.

If F_2 is not statistically significant then Fischer's ANCOVA may be employed. If F_2 is statistically significant then <u>an alternative to ANCOVA must be employed.</u>

Test 2A: Given that the covariable coefficients are the same for all groups, do the group means differ (H_1 vs H_2)? This is Fischer's Analysis of Covariance. It should be used only when F_2 is <u>not statistically significant.</u>

Calculations: $\Delta S^2 = S_1^2 - S_2^2$ where ΔS^2 is distributed

$\chi^2(k-1)$ and ΔS^2 represents the gain in sum of squares deviations due to H_1 over H_2. H_1 vs H_2 can be tested by

$$F_{2a} = \frac{\Delta S^2/(k - 1)}{S_2^2/df_{2a}}$$

where $df_{2a} = \sum_{i=1}^{k} N_i - (k + n)$.

If F_2 was significant then this test should be ignored. It has been shown that applying F_{2a} when F_2 is statistically significant consistently errodes the power of ANCOVA and generally leads to a finding of "no significant differences"[17]. If F_2 is <u>not significant</u> then one of the two following alternatives applies:

1. If F_{2a} is statistically significant, the groups differ according to the

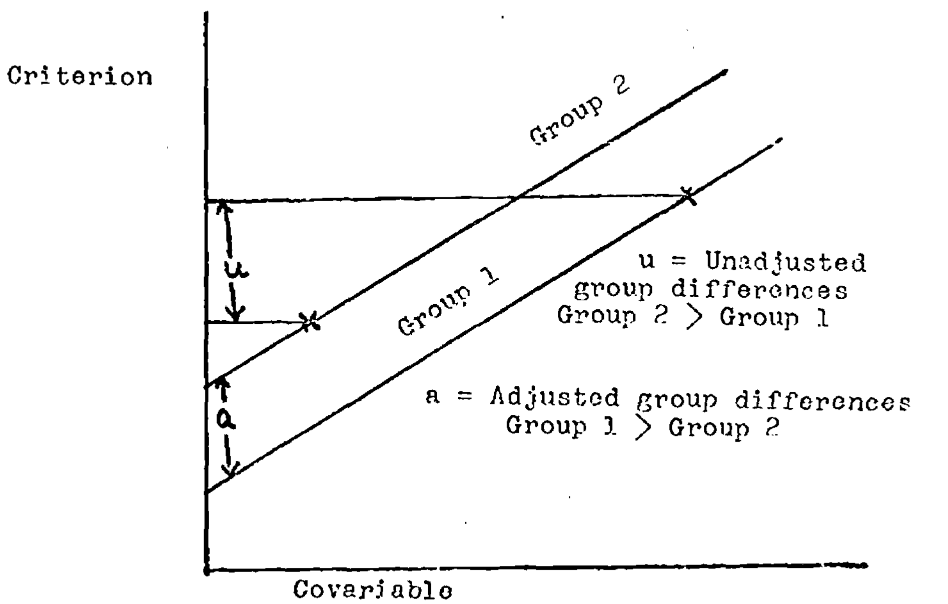
adjusted group means or group intercepts[16].

2. If F_{2a} is not statistically significant, then the adjusted group means do not differ significantly.

Figure 5 illustrates the importance of using the adjusted group menas in finally deciding how the groups differ. Although the

FIGURE 5

ADJUSTED MEANS VS ACTUAL MEANS



mean of Group 2 on the criterion is higher than that of Group 1 (by u units), it is clear that if they are equated on the covariable then Group 1 would significantly outperform Group 2 (by a units). It is impossible to

estimate the embarrassment of a researcher when he performes the lengthy and sophisticated analysis suggested above, finds significant differences, and then interpretes the relationship backwards because of a failure to identify the adjusted means.

Test 3: Are there regions where the treatments differ in effectiveness (H_j vs H_a)? (This test is only employed if F_2 is statistically significant.) This is the generalized Johnson-Neyman analysis for k groups, n covariables.

Calculations: Under H_j the maximum likelihood estimate of $\hat{\underline{B}}_i = \hat{\underline{B}}_i^a - C_i\ \underline{T}_i^{-1}\ \underline{X}'$ where $\hat{\underline{B}}_i^a$ is the estimate under H_a and $\underline{X}$ is a specified data point. $\underline{C}$ is an arbitrary set of coefficients $(C_1, C_2, \ldots, C_k)$ such that $\underline{C} = \underline{U}^{-1}\ \underline{K}$ where

$$\underline{U} = \begin{bmatrix} \underline{X}\ \underline{T}_1^{-1}\ X' - \underline{X}\ \underline{T}_2^{-1}\ \underline{X}' & 0 & \ldots & 0 & 0 \\ 0 & \underline{X}\ \underline{T}_2^{-1}\ \underline{X}' - \underline{X}\ \underline{T}_3^{-1}\ \underline{X}' & \ldots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots\ \underline{X}\ \underline{T}_{k-1}^{-1}\underline{X}' & -\ X\ \underline{T}_k^{-1}\ X' \\ 1 & 1 & 1 & \ldots\ 1 & 1 \end{bmatrix}$$

$$\underline{K} = [\underline{X}(\underline{B}_1^a - \underline{B}_2^a),\ \underline{X}(\underline{B}_2^a - \underline{B}_3^a),\ \ldots,\ \underline{X}(\underline{B}_{k-1}^a - \underline{B}_k^a),\ 0].$$

Likewise the sum of squares deviation due to H_j is

$$S_j^2 = S_a^2 + \sum_{i=1}^{k} C_i \, \underline{X} \, \underline{\hat{B}}_i^a \text{ and}$$

$$\Delta S_j^2 = S_j^2 - S_a^2 = \sum_{i=1}^{k} C_i \, \underline{X} \, \underline{\hat{B}}_i^a$$ where ΔS_j^2 is distributed χ^2 $(k - 1)$ and ΔS_j^2 represents the gain in sum of squares deviations due to H_j.

H_j vs H_a can be tested by

$$F_{j\underline{x}} = \frac{\Delta S_j^2/(k - 1)}{S_a^2/df_a}$$

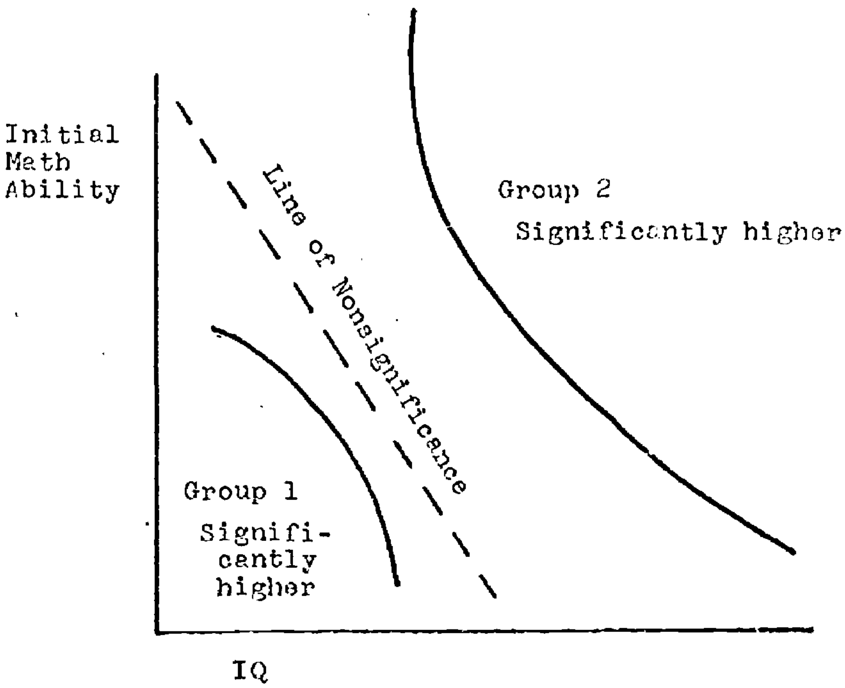
If $F_{j\underline{x}}$ is not significant then the groups do not differ significantly at point $\underline{X}$. If $F_{j\underline{x}}$ is significant then the groups differ significantly for the data point $\underline{X}$, and the products $\underline{X}\,\underline{\hat{B}}_1^a$, $\underline{X}\,\underline{\hat{B}}_2^a$, ..., $\underline{X}\,\underline{\hat{B}}_k^a$ should be examined to determine the treatment which will maximize performance for $\underline{X}$.

The results of the analysis can be used effectively to assign individuals, on the basis of covariable scores, to treatments where their predicted achievement is the highest. In the simplist situation when k = n = 2, the solution reduces to a conic section and can be graphed as illustrated in Figure 6. For this example the treatment given Group 1 is superior for students with low IQ and initial ability while the treatment given Group 2 is superior for students with high IQ and initial ability. Likewise, as new students enter the program they can be directed to the teaching method which promises the greater success based

on their initial ability and IQ.

FIGURE 6

EXAMPLE OF A JOHNSON-NEYMAN SOLUTION



Use of the Johnson-Neyman Procedure for Interaction Analysis

While the Johnson-Neyman Procedure might be considered an overly complicated substitute for ANCOVA which must be resorted to when regression slopes are not homogeneous, it can also be viewed by educators as a powerful tool for measuring the relationship between learner characteristics

and teaching straegies. There has been a great deal of interest in individual differences which affect learning performance, but generally the analyses used do not give necessary insight into the underlying dynamics. Using a computer to handle the computation, even a " non-quantitative type" can easily test, describe, and use he Johnson-Neyman methodology to classify students by their individual needs as well as by the characteristics of teaching methods. Hopefully, this kind of classification will solve some of the problems which arise because current views of educational realities are too simplistic. It may well be that rather than a nuisance in ANCOVA, treatment-covariable interaction is the key to understanding the critical relationship between teaching and learning.

Summary

The assumptions underlying covariable methods were analyzed and procedures were suggested for dealing with curvilinearity, covariable selection, and nonhomogeneity of regression. The procedure for handling nonhomogeneous group regressions was shown to be of value in assigning students to various instructional methods on the basis of their individ-

ual characteristics.

Computer programs which perform the analyses discussed in this paper are available upon request from the author. An extensive bibliography is also available.

FOOTNOTES

[1]As indicated by the title, curvilinearity and covariable effectiveness are included as ancillary topics because of their considerable impact on the quality of the analysis. For the purposes of this paper the term covariate method or covariate procedure will refer to a statistical test of group differences based on group equations of the form $Y = b_o + b_1 x_1 + \ldots + b_n x_n$.

[2]Janet D. Elashoff, "Analysis of Covariance: A Delicate Instrument", American Educational Research Journal, VI (May, 1969), 383-401, and James W. Wilson and Ray L. Cary, "Homogeneity of Regression--Its Rationale, Computation and Use," American Educational Research Journal, VI (January, 1969), 80-90.

[3]There is considerable controversy surrounding the use of covariate methods to analyze quasi-experiments, but this does not change the fact that alternative procedures are even more difficult to apply and have not been proven superior. See Donald A. Campbell and Julian Stanley, "Experimental and Quasi-experimental Designs for Research on Teaching," Handbook of Research on Teaching, ed. by N. L. Gage (Chicago: Rand McNally and Company, 1963), pp. 171-246.

[4]The discussion of these assumptions parallels Elashoff's "Analysis of Covariance."

[5]M. Atiqullah, "The Robustness of the Covariance Analysis of a One-way Classification," Biometrika, LI (December, 1964), 365-372.

[6]This is not true if the "spread" of covariable values is small or there are outlying cases. In one situation where the spread of values was small, 50 outliers caused the correlations to drop from .7 to .3 where 20,000 cases were used in the analysis. Obviously problems of this type will ruin any covariable procedure.

[7]Independence is guaranteed in a true experiment by the random assignment of subjects to groups.

[8]Campbell and Stanley, "Experimental and Quasi-Experimental Designs," pp. 231-234.

[9]Frederic Lord, "A Paradox in the Interpretation of Group Comparisons," Psychological Bulletin, LXVIII (No. 5), 304-305.

[10]Andrew Porter, The Effects of Using Fallible Variables in the Analysis of Covariance (Unpublished Ph.D. dissertation, University of Wisconsin), Ann Arbor, Michigan: University Microfilms, 1967, No. 67-12.

[11]William L. Hays, Statistics for Psychologists (New York: Holt, Rinehart and Winston, 1965), pp. 539-550.

[12]Since an infinite number of regression equations can be fitted to the data, the exact relationship must arise from a careful study of theory rather than statistical accidents.

[13]This corresponds to a singular variance-covariance matrix which cannot be inverted to complete the analysis.

[14]For example, the new variable could be the sum of the normal scores for each covariable.

[15]The derivation here is vased on Jersey Neyman's method for testing linear hypotheses as refined by Palmer O. Johnson. See Palmer O. Johnson, "The Johnsos-Neyman Technique, Its Theory and Application," Psychometrika, XV (December, 1950), 349-367; Palmer O. Johnson and Robert W. B. Jackson, Modern Statistical Methods (Chicago: Rand McNally, 1959); Palmer O. Johnson and Cyril Hoyt, "On Determining Three Dimensional Regions of Significance," Journal of Experimental Education, XV (March, 1957), 203-212; and Palmer O. Johnson and Cyril Hoyt, The Theory of Linear Hypotheses with Applications to Educational Problems (Minnesota: University of Minnesota Bureau of Educational Research, 1952).

[16]Where a variable name is underlined, it will refer to a vector or matrix of values.

[17]Percy D. Peckham, "An Inversigation of the Effects of Non-Homogeneous Regression Slopes Upon the F - Test of Analysis of Covariance," Laboratory of Educational Research Report, No. 16 (Boulder, Colorado: University of Colorado, 1960).

[18]The magnitude of the difference is identical for either although the intercepts are generally less intuitively satisfying.